

A Bayesian approach to line-transect analysis for estimating abundance

Tomoharu Eguchi*, Tim Gerrodette

Southwest Fisheries Science Center, National Marine Fisheries Service, 3333N Torrey Pines Court, La Jolla, CA 92037, USA

ARTICLE INFO

Article history:

Received 6 November 2008

Received in revised form 25 March 2009

Accepted 1 April 2009

Available online 14 May 2009

Keywords:

Line-transect

Abundance estimation

Density estimation

Half-normal detection function

Bayesian methods

Prior distribution

ABSTRACT

Line-transect analysis is a widely used method of estimating plant and animal density and abundance. A Bayesian approach to a basic line-transect analysis is developed for a half-normal detection function. We extend the model of Karunamuni and Quinn [Karunamuni, R.J., Quinn II, T.J., 1995. Bayesian estimation of animal abundance for line-transect sampling. *Biometrics* 51, 1325–1337] by including a binomial likelihood function for the number of objects detected. The method computes a joint posterior distribution on the effective strip width and the density of objects in the sampled area. Analytical and computational methods for binned and unbinned perpendicular distance data are provided. Existing information about effective strip width and density can be brought into the analysis via prior distributions. The Bayesian approach is compared to a standard line-transect analysis using both real and simulated data. Results of the Bayesian and non-Bayesian analyses are similar when there are no prior data on effective strip width or density, but the Bayesian approach performs better when such data are available from previous or related studies. Practical methods for including prior data on effective strip width and density are suggested. A numerical example shows how the Bayesian approach can provide valid estimates when the sample size is too small for the standard approach to work reliably. The proposed Bayesian approach can form the basis for developing more advanced analyses.

Published by Elsevier B.V.

1. Introduction

Estimating animal abundance is fundamental to successful management and conservation of animal populations. Line-transect analysis is a commonly used method of estimating density and abundance for a variety of species (Buckland et al., 2001). The fundamental idea behind a line-transect analysis is that the probability of detecting an object depends on the distance of the object from the transect line and, possibly, on other variables. Data are collected by moving along a predetermined transect line and recording the distances perpendicular to the transect line at which the objects of interest are detected. If the objects occur in groups or clusters, such as flocks of birds or schools of dolphins, the number of animals in each group is also recorded. The central task of a line-transect analysis is to estimate an “effective width” of the transect strip based on the observed perpendicular distances. If the effective width on each side of the transect line is μ (technically called the effective strip half-width, or ESW), the density of objects D is

$$D = \frac{\sum_{i=1}^n s_i}{2L\mu}$$

where n is the number of detected clusters (groups), s_i the number of animals in each group, and L is the total length of the transect. In a standard line-transect analysis, μ is estimated by fitting a probability density function f to the perpendicular distances, and the density estimator is

$$\hat{D} = \frac{n\hat{f}(0)\hat{E}[s]}{2L} \quad (1)$$

where $\hat{f}(0) = 1/\mu$ is the estimated probability density function of the observed perpendicular distances y evaluated at $y=0$, and $\hat{E}[s]$ is the expected cluster size (Buckland et al., 2001). Abundance N is estimated as

$$\hat{N} = A\hat{D} \quad (2)$$

where A is the total area of the study site. The estimation process in a standard line-transect analysis uses the likelihood principle because of statistically attractive properties of maximum likelihood estimators (MLE), such as consistency and invariance (e.g., Hogg and Craig, 1995; Casella and Berger, 2002).

Inference based on likelihood functions also can be made via Bayesian methods. Bayesian methods are becoming widely used in ecology because they have both theoretical and practical advantages (Ellison, 1996; Wade, 2000). Intuitively, the Bayesian approach combines existing data (prior distributions) with new data to produce an updated state of information (posterior

* Corresponding author. Tel.: +1 858 546 5615; fax: +1 858 546 7003.
E-mail address: tomo.eguchi@noaa.gov (T. Eguchi).

distributions). When new data are limited, estimation can be improved by using existing data from past or related studies. Further, interpretations of Bayesian posterior distributions are clearer than those of confidence intervals, particularly when critical decisions have to be made based on uncertain information (Goodman, 2004). Finally, hierarchical Bayesian models offer a flexible and realistic approach to ecological research (Clark and Bjørnstad, 2004; Clark et al., 2005).

Karunamuni and Quinn (1995) described a conjugate Bayesian estimation of a half-normal detection function, and showed that the Bayes estimator was better than the maximum likelihood estimator, based on relative and squared error loss. We extend this work to full density estimation by including a binomial likelihood function for the number of objects detected on the transect line, which has been discussed in the non-Bayesian framework (e.g., Seber, 1973; Sen et al., 1974; Quinn and Gallucci, 1980; Borchers and Burnham, 2004). We formulate prior information on the detection function in terms of the ESW (μ), a central concept in line-transect analysis (Burnham et al., 1980). Perpendicular distances may either be recorded exactly (“unbinned” data) or recorded less precisely in distance categories (“binned” data). The performance of the proposed Bayesian approach is tested with previously analyzed line-transect data sets as well as simulated data. We compare the Bayesian approach to results using the standard line-transect software Distance (Thomas et al., 2003) by assessing bias and precision with simulated data. The proposed approach does not incorporate advanced developments of line-transect analyses (Buckland et al., 2004). However, the proposed approach can be extended to more advanced settings.

2. Methods

In the following sections, we introduce notation and assumptions (Section 2.1), explain the overall approach (Section 2.2), describe likelihood functions for the number of detected objects (Section 2.3.1), unbinned perpendicular distances (Section 2.3.2) and binned perpendicular distances (Section 2.3.3), illustrate prior distributions for ESW (Section 2.4.1) and density (Section 2.4.2), describe the method of obtaining posterior distributions (Section 2.5), and introduce performance testing processes (Section 2.6).

2.1. Notation and assumptions

Following Buckland et al. (2001), we use standard line-transect notation (Table 1). The standard assumptions of a line-transect analysis are (e.g., Buckland et al., 2001):

1. The population of interest is closed from immigration, emigration, birth, and death during the survey.

Table 1
Notation.

Symbol	Description
A	Total area of the study site
L	Length of transect
W	Maximum perpendicular distance of detected objects, or truncation distance
a	Sampled area = $2LW$
n	Number of objects detected within sampled area a
\mathbf{y}	Perpendicular distances of the n detected objects
N_a	Number of objects within sampled area a
P_a	Probability of detecting an object within sampled area a
$f(y)$	Probability density function of perpendicular distances
$g(y)$	Probability of detection as a function of perpendicular distance
θ	Parameter(s) of the detection function $g(y)$
μ	Effective strip half-width (ESW)
λ	Precision (1/variance) for the half-normal detection function

2. The transect line is placed randomly in the study area such that sampled area (a) is a random sample of the total area (A).
3. Objects do not move in response to the observation process before they are detected.
4. Measurements of perpendicular distance are exact.
5. Distance data are independent of the total number of objects.
6. Objects on the track line are detected with certainty; that is, $g(0) = 1$.
7. The detection function $g(y)$ is proportional to the probability density function $f(y)$ for perpendicular sighting distances y ; that is, $g(y) \propto f(y)$ for all y .
8. Detection of each object is an independent event, regardless of the spatial distribution of the objects.

Note that assumptions (6) and (7) together imply that $\mu = 1/f(0)$ and $g(y) = \mu f(y)$.

2.2. Overview

In this initial development of a basic Bayesian model, we assume that objects occur singly (i.e., $E[s] = 1$ in Eq. (1)) and that probability of detection depends on perpendicular distance only. Following standard practice, we assume A , L , W and a are known without error, but that data n and \mathbf{y} are observed as a result of stochastic processes. The parameter of primary interest to be estimated is N_a , the number of objects in the sampled area, from which density and total abundance can be estimated. The overall approach to the following development is to compute the conditional joint probability distribution of N_a and θ , which includes all parameters that affect $f(y)$. According to Bayes' theorem, and making use of the assumed independence of N_a and θ (assumption (5)), the joint posterior is:

$$p(N_a, \theta | \mathbf{y}, n) \propto p(n | N_a, \theta) p(\mathbf{y} | \theta) p(N_a) p(\theta), \quad (3)$$

where $p(n | N_a, \theta)$ is the likelihood function for n , $p(\mathbf{y} | \theta)$ is the likelihood function for \mathbf{y} , and $p(N_a)$ and $p(\theta)$ are the prior distributions for N_a and θ , respectively. The marginal posterior distribution of N_a is obtained from (3) by integrating over θ :

$$p(N_a | \mathbf{y}, n) = \int_{-\infty}^{\infty} p(N_a, \theta | \mathbf{y}, n) d\theta, \quad (4)$$

where the integral may be over more than one parameter. Using the conventional design-based approach, density is estimated as $D = N_a/a$ and total abundance as $N = DA$ (Eq. (2)), because both a and A are assumed known without error. Existing information about abundance and the detection process is included in analysis via the prior distribution on abundance and the prior distribution on the parameters of the detection function. If no previous information exists, either or both of these functions may be set to uniform or other distributions that provide equal or nearly equal probabilities over a broad range of possible values. We call such distributions vague prior distributions. Although these distributions also have been called noninformative or uniform distributions, we prefer “vague prior” because these distributions can be informative and they may not be strictly uniform.

2.3. Likelihood functions

2.3.1. Number of detections

In (3), we first consider the likelihood function $p(n | N_a, \theta)$ for the number of objects n detected in the sampled area. It is convenient to define P_a , the probability of detecting an object within the sampled area a as

$$P_a(\theta) = \frac{\int_0^W g(y | \theta) dy}{W} = \frac{\mu}{W}. \quad (5)$$

The notation emphasizes that the probability of detecting an object within the sampled area is a function of the parameters θ of the detection function. Geometrically, detection probability is the proportional area under the detection function within the $W \times g(0)$ rectangle (Buckland et al., 2001). For notational brevity, we now write P_a instead of $P_a(\theta)$.

Using (5) and assumption (8), each sighting is modeled as a Bernoulli trial with common detection probability P_a (Seber, 1973; Borchers and Burnham, 2004). Given the total number of available objects in the sampled area N_a and the detection probability P_a , the observed number of objects n follows a binomial distribution:

$$p(n|N_a, \theta) = \binom{N_a}{n} P_a^n (1 - P_a)^{N_a - n}. \quad (6)$$

Note that Borchers and Burnham (2004) used N_c (our N_a) to denote the total number of available objects in the sampled area.

2.3.2. Unbinned perpendicular distances

In (3), we next consider the likelihood function $p(\mathbf{y}|\theta)$ for perpendicular distances, in the case when distances are measured exactly (unbinned data). The half-normal detection model has a single parameter, the variance σ^2 . Following Karunamuni and Quinn (1995), it is more convenient to work with precision $\lambda = 1/\sigma^2$. For the half-normal detection function, λ is directly related to ESW (μ) and $f(0)$ as $\lambda = \pi/(2\mu^2) = \pi/(2f(0)^2)$. We assume that knowledge of λ can be represented by a gamma distribution with hyperparameters α and β , where $\lambda > 0$, $\alpha > 0$, and $\beta > 0$. Because of the conjugate property between normal and gamma distributions, the joint posterior of N_a and λ conditional on the observed data is the following (Appendix A):

$$p(N_a, \lambda|\mathbf{y}, n) \propto p(N_a) \lambda^{\alpha_2 - 1} \exp(-\lambda \beta_2) \frac{\Gamma(N_a + 1)}{\Gamma(N_a - n + 1)} P_a^n (1 - P_a)^{N_a - n}, \quad (7)$$

where $\alpha_2 = \alpha + n/2$ and $\beta_2 = \beta + \sum_{i=1}^n y_i^2/2$, $i = 1, \dots, n$.

2.3.3. Binned perpendicular distances

If, instead of being recorded exactly, perpendicular distances \mathbf{y} are grouped into u groups or bins, the likelihood function is multinomial with bin-specific probabilities (Buckland et al., 2001, p. 62):

$$\omega_i = \int_{c_{i1}}^{c_{i2}} f(y|\theta) dy, \quad (8)$$

where (c_{i1}, c_{i2}) is the interval spanning the i -th group, $i = 1, \dots, u$. The likelihood function of the parameter is:

$$L(\theta|\mathbf{n}, \omega) = \frac{n!}{\prod_{i=1}^u n_i!} \prod_{i=1}^u \omega_i^{n_i} \quad (9)$$

where $n = \sum_{i=1}^u n_i$, $\mathbf{n} = [n_1, \dots, n_u]$, and $\omega = [\omega_1, \dots, \omega_u]$.

For the half-normal detection function $\theta = \{\lambda\}$, as in the unbinned data. Combining (8) and (9), the probability function of data conditional on parameters θ is:

$$p(\mathbf{y}|\theta, \mathbf{n}) = \frac{n!}{\prod_{i=1}^u n_i!} \prod_{i=1}^u \left(\int_{c_{i1}}^{c_{i2}} f(y_i|\theta) dy \right)^{n_i}. \quad (10)$$

This likelihood function (10) is used in (3) to obtain the posterior distribution for N_a and λ .

2.4. Prior distributions

Next we consider the prior distributions $p(\theta)$ and $p(N_a)$ in Eq. (3). In the case of the half-normal detection function, $p(\theta) = p(\lambda)$. If prior

information on N_a or λ exists, it will usually be in terms of density D and effective strip width μ . Therefore, we use simple algebraic relationships between λ and effective strip width ($\lambda = \pi/(2\mu^2)$) and between N_a and density ($N_a = aD$) to develop the priors on λ and N_a .

2.4.1. Effective strip width

Prior information on ESW (μ) needs to be transformed into a gamma prior distribution on λ to utilize the conjugate approach described in (Section 2.3.2). We use a numerical approach to accomplish this task. Given an estimated mean $\hat{\mu}$ and variance v_μ of ESW from auxiliary data, we find appropriate gamma parameters $\alpha_\mu = \hat{\mu}^2/v_\mu$ and $\beta_\mu = \hat{\mu}/v_\mu$. We then generate random variables μ_r from this gamma distribution and fit a second gamma distribution to the values of $\pi/(2\mu_r^2) (= \lambda_r)$ using the method of moments (Appendix B). Although unsophisticated, this approach provides a practical and adequate method for including prior data on ESW into the conjugate Bayesian analysis (Appendix C). Alternatively, a non-conjugate approach could represent prior information on μ directly with, for example, a gamma distribution.

2.4.2. Density

Prior information about density D may be represented by a gamma distribution:

$$p(D|\gamma, \kappa) = \frac{\kappa^\gamma}{\Gamma(\gamma)} D^{\gamma-1} \exp(-D\kappa), \quad (11)$$

where $D > 0$, $\gamma > 0$, and $\kappa > 0$. A non-vague prior distribution may be constructed from a previous estimate of density \hat{D} with variance v_D . The parameters of a prior gamma distribution for D can be obtained by using $\gamma = \hat{D}^2/v_D$ and $\kappa = \hat{D}/v_D$. The prior distribution $p(N_a)$ is a simple transformation of $p(D|\gamma, \kappa)$ because $D = N_a/a$. Consequently, $p(N_a)$ is also a gamma distribution with parameters $(\gamma, \kappa/a)$. When using the gamma prior distribution for the density, the joint posterior for unbinned data becomes:

$$p(N_a, \lambda|\mathbf{y}, n) \propto N_a^{\gamma-1} \lambda^{\alpha_2-1} \times \exp\left(-\lambda \beta_2 - \frac{\kappa N_a}{a}\right) \frac{\Gamma(N_a + 1)}{\Gamma(N_a - n + 1)} P_a^n (1 - P_a)^{N_a - n}, \quad (12)$$

because of the conjugate property between two independent gamma distributions.

In practice, a prior distribution for a previously studied area may be constructed from the mean and variance estimates of the density (\hat{D} and v_D , respectively). For areas with no previous estimates, the gamma parameters (γ and κ) can be set such that it provides a vague prior distribution. Small values of γ and κ , or equivalently large variance of the density (v_D) with respect to the mean (\hat{D}), result in such distributions.

2.5. Posterior inference

We use a Markov chain Monte Carlo (MCMC) method, specifically the Metropolis-Hastings algorithm, to sample from the joint posterior distribution of N_a and λ . Details of the algorithm can be found in Metropolis et al. (1953) and Hastings (1970). Less technical references on the algorithm and other MCMC methods include Martinez and Martinez (2002), Berg (2004), and Clark (2007). The Metropolis-Hastings algorithm produces random samples from a distribution by using the first order Markov process. To implement the algorithm, computer programs were written in the programming language C, whereas Matlab (The MathWorks, Inc., Natick, MA) was used for user interface, data manipulation, and graphical outputs. We used two phases of the algorithm to tune the

computation process. During the first phase (burn-in phase), a wide range of the parameter space was searched for a high-density area by running multiple independent Markov processes (hereafter called Markov chains). The starting point of each Markov chain was randomly selected from uniform distributions for the parameters. Resulting posterior samples from the burn-in phase were used to make the algorithm more efficient for the second phase (sampling phase). For our analysis, we used four independent Markov chains of 10,000 steps as a burn-in phase, followed by four independent Markov chains of 100,000 steps as a sampling phase. To avoid the serial correlations within each chain during the sampling phase, every third step was retained for computing the summary statistics. Samples from the burn-in phase were not used for the inference process. Convergence of Markov chains was determined by using the \hat{R} statistic (Gelman et al., 2004). The algorithm for unbinned data also was implemented in WinBugs (Lunn et al., 2000; results not shown). The WinBugs codes are in Appendix D, whereas C and Matlab programs are available on request from the first author.

2.6. Performance testing

To test the performance of the proposed method, we analyzed four datasets with the proposed Bayesian approach. The same datasets were analyzed with Distance software, which uses a combination of maximum likelihood and Horwitz–Thompson estimators (Thomas et al., 2003). We analyzed two real datasets which are publicly available, one binned and one unbinned, and two simulated datasets which are available from the first author. For the simulated data, performance was measured by precision (widths of 95% posterior (Bayesian) and 95% confidence (Distance) intervals) and bias. For both density and ESW, bias was computed as the difference between the point estimate (median of posterior for Bayesian) and the true value. For some examples, we also computed the mode of the posterior by smoothing with a normal kernel density function and then numerically differentiating the smoothed posterior.

2.6.1. Real data

The real unbinned dataset was collected by Laake (1978), who placed 150 wooden stakes randomly within a rectangular area of sagebrush-grass near Logan, Utah. The length of the transect line (L) was 1000 m. The true density of the stakes was 37.5 stakes per hectare. The dataset has been used in a variety of literature (e.g., Burnham et al., 1980, p. 61, Quang, 1990) and is available on the Internet. In Distance, we used the half-normal distribution as the key detection function, and no adjustment terms were considered. We analyzed the data with perpendicular distances truncated at 20 m and also without truncation. For the Bayesian analysis, we used a vague prior distribution for the density within the study area. For the prior distribution of ESW, we used a gamma distribution with a mean of 10 m and variances of 0.02, 0.2, or 2.0 (Fig. 1A). The small variance provides a prior for which the majority of density is concentrated over a small range of ESW, whereas the large variance is for a vague prior.

The real binned data are line-transect counts of duck nests of unknown species at the Monte Vista National Wildlife Refuge in Colorado during 1969–1974 and 1986–1987 (Buckland et al., 2001, p. 341). Details of the data collection and effort can be found in Gilbert et al. (1996). The true density of duck nests is unknown in this case. The total effort was $L = 1244.3$ miles (2002.5 km), and the maximum observed distance was $W = 11.5$ ft (3.5 m). As with the wooden stake data, we used half-normal model with no adjustment terms for Distance. A vague prior for density and a range of priors were used for ESW (Fig. 1B). We purposely set the prior mean at

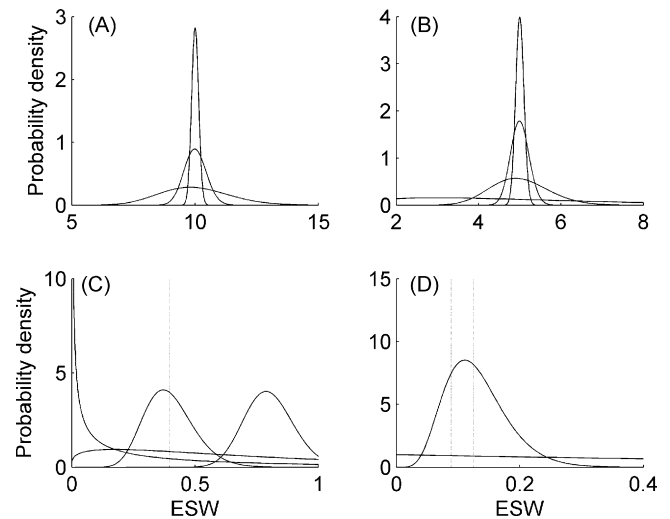


Fig. 1. Prior distributions for effective strip width (ESW) used in the analyses. (A) The three prior distributions used for wooden stake data analysis. (B) The four prior distributions used for duck-nest data analysis. (C) The four prior distributions used for the analysis of 500 simulated data. The vertical dotted line indicates the true ESW (0.396 km). (D) The two prior distributions for the simulated two-year dataset. The dotted vertical lines indicate the true ESW for two years (0.125 km and 0.089 km).

5 ft (1.5 m), which was less than the estimate from Distance, to show the effects of inaccurate prior distributions on the posterior distributions.

2.6.2. Simulated data

To simulate line-transect data, we used the method described in Buckland et al. (2001) with the half-normal detection function. Expected sample size ($E[n]$) of each simulated dataset was a function of the length of the transect line (L), the variance of the half-normal detection function (σ^2), and the true density (D) of the objects (Buckland et al., 2001, pp. 92–93).

$$E[n] = \frac{2 \times L \times D}{f(0)} \quad (15)$$

where $f(0) = \sqrt{2/(\pi\sigma^2)}$. Sample size for each simulation was randomly drawn from a Poisson distribution with the expected sample size. The true density was fixed at 0.1 km^{-2} . The true detection function was half-normal with variance $\sigma^2 = 0.1$, which corresponded to a true ESW of 0.396 km. A total of 500 simulated datasets were analyzed as ‘unbinned’ and ‘binned’ data. Binned data were created using Distance, where 10 bins of equal width were used.

To assess the effects of prior distributions on the analysis, we analyzed the simulated data with four different priors on ESW (μ) from two means (0.396 and 0.8 km) and two variances (0.01 and 0.5; Fig. 1C). The prior with $\hat{\mu} = 0.396$ km and $v_{\mu} = 0.5$ was an exponentially decreasing function with respect to ESW, whereas the prior with $\hat{\mu} = 0.8$ km and $v_{\mu} = 0.5$ provided essentially uniform probability over the parameter space of interest (Fig. 1C).

Finally, to demonstrate the ability of the Bayesian method to use prior information to improve estimates of abundance with scarce data, we simulated a two-year study, in which the density was unchanged over the two years, but ESW and effort (L) changed. This simulates a common situation where weather or logistical conditions prohibit researchers from allocating similar effort in two consecutive years. The first year’s data were simulated with the true density of $D_1 = 0.04 \text{ km}^{-2}$, the true vari-

ance of the detection function was $\sigma^2_1=0.01$ ($\mu_1=0.125$ km), and the effort was 5000 km. For the second year, the underlying true density did not change, $D_2=0.04$ km⁻², but the effective strip width decreased; $\sigma^2_2=0.005$ ($\mu_2=0.089$ km), and the effort decreased by 80% (1000 km). We analyzed the first dataset with vague prior distributions on ESW ($\hat{\mu}=1.0$, $v_\mu=1.0$; Fig. 1D) and a vague prior distribution on density. The result from the analysis for the first year was used for the prior distributions in the analysis of the second year's data. The same datasets also were analyzed using Distance, treating them as two independent datasets and with a hierarchical approach using the data for both years to assume a common detection function for both years.

3. Results

3.1. Real data

Despite inclusion of the additional uncertainty of the sampling process in the model, i.e., the binomial term, the proposed Bayesian approach performed well in all performance tests. A precise posterior distribution was obtained for the wooden stake data (Fig. 2) and convergence of MCMC was reached according to the \hat{R} statistic. There was a negative correlation between density and ESW in the joint posterior distribution (Fig. 2). The variance of prior distributions had small effects on the posterior distributions (Table 2). The median density was approximately 34 stakes per hectare, where the widths of 95% posterior probability intervals (PI) were <17 (Table 2). The estimated density by Distance was 33.1, whereas the width of the 95% confidence interval (CI) was approximately 18 (Table 2). The true density (37.5 stakes per hectare) was within posterior and confidence intervals. Truncation of data at 20 m resulted in narrower ESW and greater density for both Distance and the Bayesian approach (Table 2). The Bayesian approach resulted in narrower intervals than results of Distance (2.5 m vs. 3.1 m and 16.6 ha⁻² vs. 21.3 ha⁻²; Table 2).

For the duck-nest data, the estimate of ESW from Distance was 7.6 ft (2.3 m) and the estimated density was 58.2 nests per square mile (95% CI=48.5–69.8; Table 3). The Bayesian approach indicated the obvious consequences of precise but inaccurate prior distributions on inference (Fig. 3). For the extreme case, where the prior mean on ESW was 5.0 ft and the variance 0.01, the posterior mean for the ESW was less than the Distance estimate and the density was greater than the Distance estimate (Table 3). The difference decreased as the variance of the prior distribution increased (Table 3). When the prior variance was 10 (i.e., vague prior), the medians of the posterior were approximately equal to the estimate from Distance, as expected. The widths of PI's were narrower than those of the CI's (Table 3).

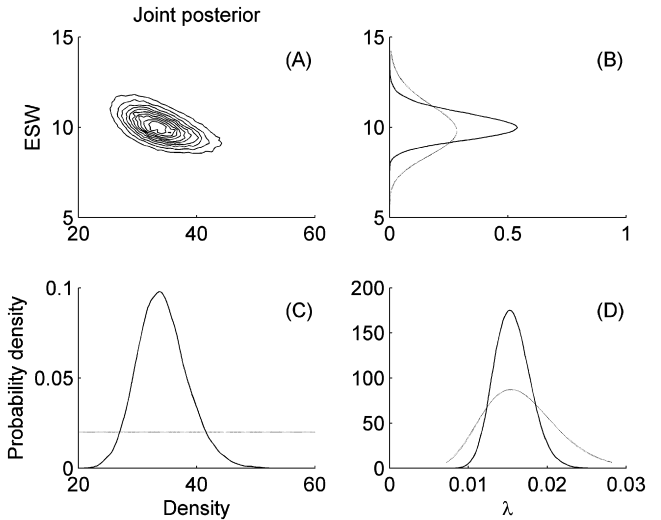


Fig. 2. Joint and marginal posterior distributions of effective strip width (ESW) and density for the analysis of wooden stake data. The contour plot (A) is the joint posterior distribution, whereas (B) and (C) are the marginal distributions. Prior distributions are shown as dotted lines. Only one set of prior and posterior distributions is shown. The prior distribution for ESW had the $\hat{\mu}=10$ and $v_\mu=2.0$, where as a vague prior was used for density. The prior and marginal posterior on the precision (λ) are shown in (D). The prior for λ was calculated from the prior for ESW using the method described in the text.

3.2. Simulated data

Analyses of simulated datasets allowed us to evaluate effects of sample size and prior distributions on bias and precision of posterior distributions. Convergence was reached for all MCMC implementations according to the \hat{R} statistic. For unbinned datasets, the inclusion proportion of 95% PI's for ESW was approximately 0.97, whereas for 95% CI's it was 0.41. Widths of 95% PI's were generally narrower than those of 95% CI's regardless of the choice of prior distribution (Fig. 4). In general, widths of 95% PI's and CI's had long tails, where CI's had more extreme values than the PI's. Similar results were found for when data were binned (results not shown). Effects of priors were less for the binned data analysis than for the unbinned datasets. For binned datasets, the inclusion proportion of 95% PI's for ESW was approximately 0.97, whereas that of 95% CI's was 0.46.

For unbinned and binned datasets, biases of density for simulated datasets indicated the Bayesian and the Distance approaches provide similar results, except when the prior was inaccurate and precise ($\hat{\mu}=0.8$, $v_\mu=0.01$; Fig. 5). Such an informative prior overwhelmed the data, and the estimated density was negatively biased. For the same prior, the ESW was estimated consistently greater

Table 2
Summary statistics of analyses of wooden stake data. Prior mean and variance of ESW are indicated by $\hat{\mu}$ and v_μ , respectively. D indicates density and μ indicates ESW. Median (Med), mode, and 95% posterior intervals are presented for Bayesian analysis. For the maximum likelihood approach, the point estimate (Distance) is presented under Mode and 95% CI are shown for the interval. Rows with "Truncation" headings indicate results for analyses when perpendicular distances were truncated at 20 m.

	Prior		μ			D		
	$\hat{\mu}$	v_μ	Med	Mode	Interval	Med	Mode	Interval
Bayesian	10.00	0.02	10.0	10.0	[9.7, 10.3]	34.2	34.3	[28.0, 41.4]
	10.00	0.20	10.0	10.0	[9.3, 10.9]	34.2	34.3	[27.5, 41.9]
	10.00	2.00	10.1	10.1	[8.8, 11.8]	33.6	34.3	[26.4, 42.4]
Truncation	10.00	2.00	9.2	9.2	[8.1, 10.6]	36.6	36.4	[29.7, 46.3]
Distance				10.3	[9.1, 11.6]		33.1	[25.3, 43.3]
Truncation				9.4	[8.0, 11.1]		35.7	[26.6, 47.9]

Table 3

Summary statistics of analyses of binned duck-nest data. Prior mean and variance of ESW are indicated by $\hat{\mu}$ and v_{μ} , respectively. D indicates density and μ indicates ESW. Data were grouped into 11 equal-width bins. Posterior distributions are summarized in median (Med), mode, and 95% posterior interval. For the maximum likelihood approach, the point estimate (Distance) is presented under Mode and 95% CI are shown in the interval column.

	Prior		μ			D		
	$\hat{\mu}$	v_{μ}	Med	Mode	Interval	Med	Mode	Interval
Bayesian	5.0	0.01	5.3	5.3	[5.10, 5.47]	83.2	81.9	[74.8, 92.7]
	5.0	0.05	6.0	5.9	[5.59, 6.35]	74.0	74.3	[66.0, 82.3]
	5.0	0.5	7.1	7.2	[6.41, 7.90]	61.8	59.4	[54.1, 71.0]
	5.0	10.0	7.5	7.5	[6.73, 8.54]	58.3	58.6	[50.5, 67.5]
Distance				7.6	[6.69, 8.51]		58.2	[48.5, 69.8]

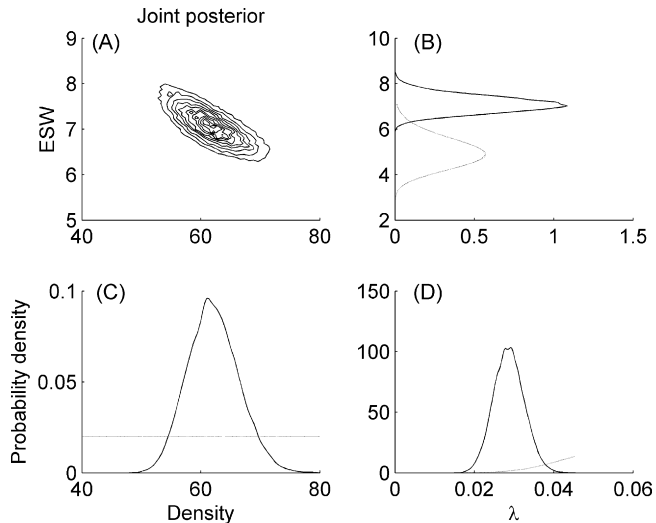


Fig. 3. Joint and marginal posterior distributions of effective strip width (ESW) and density for the analysis of duck-nest data. The contour plot (A) is the joint posterior distribution, whereas (B) and (C) are the marginal distributions. Prior distributions are shown as dotted lines. Only one set of prior and posterior distributions is shown, where $\hat{\mu} = 5$ and $v_{\mu} = 0.5$ and a vague prior on density. The prior and marginal posterior on the inverse of variance (λ) are shown in (D). The prior for λ was calculated from the prior for ESW using the method described in the text. The abscissa of the bottom right figure is truncated to show the width of the posterior distribution.

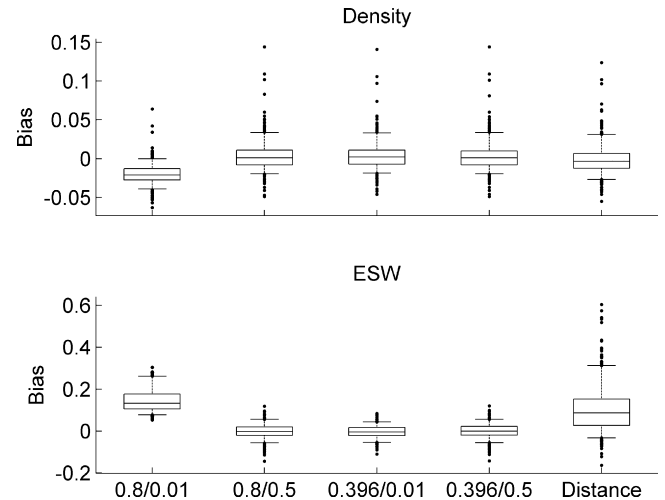


Fig. 5. Distribution of bias of density and effective strip width (ESW) for analyses of 500 simulated datasets when data were not binned. Boxes show inter-quartile ranges (25–75 percentiles) and whiskers 90-quantile ranges (5–95 percentiles). The horizontal line in each box is the median. The prior mean and variance of ESW are indicated along the abscissa ($\hat{\mu}/v_{\mu}$). Distance indicates the analysis using Distance software.

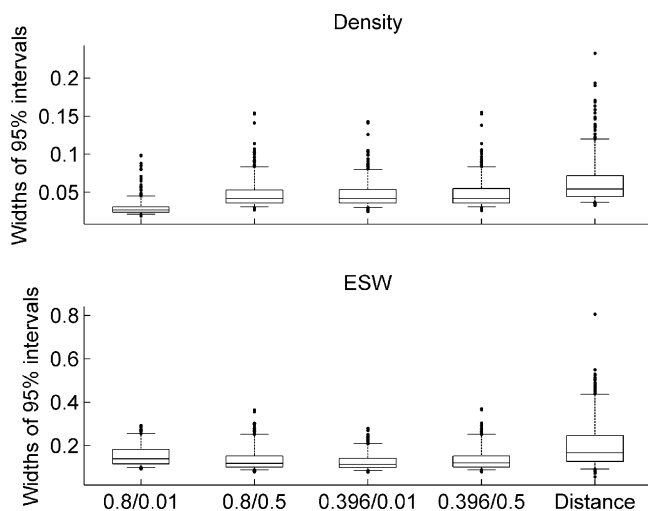


Fig. 4. Distribution of widths of 95% posterior and confidence intervals for estimates of density and ESW, for analyses of 500 simulated datasets when data were not binned. Boxes show inter-quartile ranges (25–75 percentiles) and whiskers 90-quantile ranges (5–95 percentiles). The horizontal line in each box is the median. The prior mean and variance of ESW for the Bayesian analyses are indicated along the abscissa ($\hat{\mu}/v_{\mu}$). Distance indicates the analysis using Distance software.

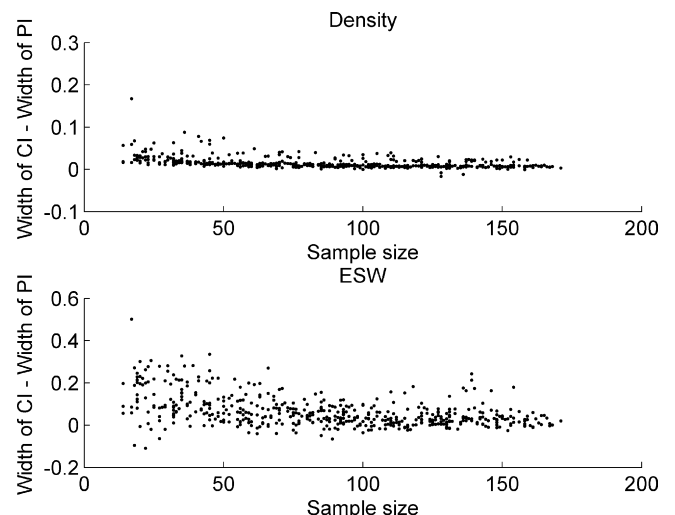


Fig. 6. The difference in widths of 95% confidence (CI) and posterior (PI) intervals for density and effective strip width (ESW), as a function of sample size (i.e., the number of detected objects) for the analysis of 500 simulated datasets. Data were not binned. Results of Bayesian analyses were based on the prior distribution with $\hat{\mu} = 0.8$ and $v_{\mu} = 0.5$, and a vague prior on density.

Table 4

Point and 95% interval estimates for the analysis of the simulated two-year dataset. D indicates density and μ indicates ESW. The true ESW were 0.125 km for the first year and 0.089 km for the second year. True densities were 0.04 km^{-2} for the both years. The prior mean and variance of ESW for the first year were, $\hat{\mu} = 1.0$, and $v_{\mu} = 1.0$, whereas for density, a vague prior was used. For the second year, the prior means and variances were $\hat{\mu} = 0.12$, $v_{\mu} = 0.004$, and $\hat{D} = 0.04$, $v_D = 0.0001$. Estimates for the second year using Distance (Distance¹) were given with a warning (see text for details). A global model for the two years was used to alleviate the small sample size of the second year in Distance². The point estimates for Bayesian analysis were medians of posterior distributions.

	Bayesian	Distance ¹	Distance ²
Year 1			
μ	0.12 [0.10, 0.15]	0.13 [0.11, 0.16]	0.12 [0.10, 0.15]
D	0.05 [0.04, 0.06]	0.04 [0.03, 0.06]	0.05 [0.03, 0.06]
Year 2			
μ	0.08 [0.06, 0.13]	0.09 [0.05, 0.15]	0.12 [0.10, 0.15]
D	0.04 [0.03, 0.06]	0.05 [0.02, 0.12]	0.03 [0.02, 0.07]

than the true ESW. The bias for ESW was greater for the Distance approach than for the Bayesian approach.

For both binned and unbinned data, we found that the Bayesian approach performed better, in terms of widths of 95% intervals, than Distance when the sample size was small (Fig. 6). The difference was insignificant when the sample size was greater than approximately 100. For the majority of the analyses of simulated datasets, PI's were narrower than CI's (Fig. 6).

Analyses of the simulated two-year data set illustrated how the Bayesian method may be used when information is available from a previous year. Using Distance, the point estimate of the ESW for the first year was 0.13 km (95% CI = 0.11–0.16), whereas for density, the point estimate was 0.04 km^{-2} (95% CI = 0.03–0.06). Using vague prior distributions, the median of the marginal posterior distribution for ESW was 0.12 km (95% PI = 0.10–0.15), whereas for density, it was 0.05 km^{-2} (95% PI = 0.04–0.06; Table 4). The true density for the first year was 0.04 km^{-2} whereas the true ESW was 0.125 km.

For the second year, because of much less effort ($L = 1000 \text{ km}$) and a narrower ESW ($\mu = 0.089 \text{ km}$), the sample size was only eight. Using solely the second year's data the point estimate from Distance

for the ESW was 0.09 km (95% CI = 0.05–0.15), whereas for the density, it was 0.05 km^{-2} (95% CI = 0.02–0.12 km^{-2}). However, Distance software provided a warning about the small sample size, cautioning not to expect reasonable results. In other words, although the estimates were close to the real values, no reliable estimate was available using the Distance approach. A hierarchical approach in Distance alleviated the problem of small sample size, i.e., a global model for all data and separate estimates of density for two-year datasets. Distance reported the point estimate of ESW to be 0.12 km (95% CI = 0.10–0.15), whereas for densities, the point estimates were 0.05 km^{-2} (95% CI = 0.03–0.06 km^{-2}) for the first year and 0.03 km^{-2} (95% CI = 0.02–0.07 km^{-2}) for the second year. For the Bayesian analysis, we used the results from the analysis for the first year's data to build the prior distribution for the second year's data, assuming the similar sampling process ($\hat{\mu} = 0.13 \text{ km}$ and $v_{\mu} = 0.05^2$). For density, we used $\hat{D} = 0.05$ and $v_D = 0.02^2$. Convergence was reached for the MCMC according to the \hat{R} statistic. The median of the ESW was 0.08 km (95% PI = 0.06–0.13 km), whereas for density, it was 0.04 km^{-2} (95% PI = 0.03–0.06; Fig. 7).

4. Discussion

The analysis of wooden stake and duck-nest datasets demonstrated the utility of the proposed Bayesian approach. Although the differences in point estimates between the Bayesian and Distance approaches were small, widths of uncertainty measures (CI's and PI's) were generally less for the Bayesian approach. With these datasets, the effects of prior distributions were negligible, indicating even a moderate sample size ($n = 68$ in the wooden stake dataset) can override the influence of prior distributions. Truncation of data also resulted in more precise estimates for the Bayesian approach than for the Distance approach. An additional analysis on the wooden stake dataset with a more vague prior distribution ($v_{\mu} = 50$) resulted in narrower 95% PI for ESW (7.7–10.6) and for density (29.9–47.9) than the results of Distance (results not shown).

Prior distributions are a unique feature of Bayesian statistics. Prior distributions bring data from past or related studies to the current analysis. From the Bayesian point of view, the current state of knowledge is a combination of what was previously known and what new data indicate. Analysis of simulated data provided a comprehensive examination of the performance of the Bayesian approach with respect to sample size and prior distributions. Because true values were known, bias could be assessed. The Bayesian analysis generally gave less biased estimates of ESW (μ), while both the Bayesian method and Distance gave unbiased estimates of density (Fig. 5). Distance performed better than the Bayesian method when the prior distribution in a Bayesian analysis was “wrong” in the sense that it was different from the current data. In terms of precision, confidence intervals of Distance were slightly larger than probability intervals (Fig. 6). The tendency for confidence intervals to be larger than probability intervals decreased with sample size, for estimates of ESW and density (Fig. 6).

We have shown through the analysis of the simulated two-year dataset that the informative prior distribution can improve the precision of the posterior distribution, especially when the sample size is small. Financial and logistical constraints may often prohibit researchers from obtaining an ideal sample size. By using the Bayesian approach, one can maximize usable historic information about the system. Although the Distance approach can use auxiliary data to construct a global detection function for a small dataset such that estimation can be accomplished, this approach makes an implicit assumption that the detection function was

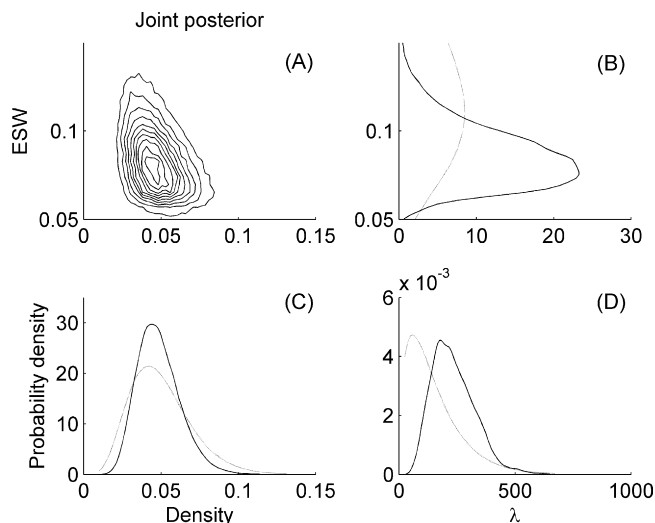


Fig. 7. Posterior distributions for the analysis of the second year of the simulated two-year dataset. Prior distributions (dotted lines) were constructed from the results of the analysis of the first year's data. The contour plot (A) is the joint posterior distribution, whereas (B) and (C) are the marginal distributions. The prior and marginal posterior on the precision (λ) are shown in (D).

identical for the two datasets. The assumption, however, is often not valid. The Bayesian approach does not assume equal detection functions, but uses the information from the ancillary data in the form of prior distributions to facilitate the inference for the small dataset.

The choice of prior distributions has been a contentious issue between the anti-Bayesians and Bayesians (e.g., Dennis, 1996, 2004; Gelman, 2008a,b). In our view, a prior distribution is not a measure of one's belief but a probabilistic measure of knowledge constructed from available information. If one wishes the current analysis to be independent of previous information, vague prior distributions can be used to force the posterior distribution to be solely a function of likelihood functions and new data. On the other hand, if one wishes to include existing data that are relevant to the current analysis, non-vague prior distributions provide a means of doing that. The process of creating a prior distribution requires comprehensive examinations of the existing information on the parameters, which facilitates a thorough understanding of the system, including the sampling and data-generating processes (Goodman, 2004). We, therefore, suggest using various defensible priors in an analysis and presenting results for all priors under consideration. For a large sample size, the choice of priors may be inconsequential. For a small sample size, the choice of prior should be explained.

The proposed Bayesian approach to line-transect analysis can be extended in several ways. First, objects could occur in groups rather than singly. Second, detection on the transect line might not be perfect so that $g(0) < 1$. If the data estimating $g(0)$ were based on a separate study, the results of Raftery and Schweder (1993) on a Bayesian approach to inference on the ratio of two independent parameters would be relevant. Third, other detection functions could be used (Buckland, 1985), particularly the two-parameter hazard-rate function. Adjustment terms to a detection function could be used for additional flexibility to fit the detection function (Buckland et al., 2001). Such adjustments result in detection functions that are not proper probability density functions, but they could be standardized numerically in a Bayesian analysis. Fourth, covariates that affect detection probability could be included (Marques and Buckland, 2004; Gerrodette and Forcada, 2005). Fifth, the assumption of independence of sightings could be relaxed by considering likelihood functions other than the binomial for the number of detections (Sen et al., 1974). Sixth, the approach can be treated as a component of a larger hierarchical model, where other processes and states may be included, such as different age groups, locations, movements, and sexes (Thomas et al., 2004; Carlin et al., 2006).

In this development, we did not consider the variability associated with the coverage probability (a/A). We treated the sampled area ($a = 2LW$) as a fixed quantity, as it is done in a conventional distance analysis. In reality, however, the coverage probability is rarely equal throughout the study region (Borchers and Burnham, 2004). Strindberg et al. (2004) introduced an analytical method to deal with variable coverage probabilities for computing the total abundance. In our development, a posterior distribution of N_a , hence D , can be obtained for each transect line. To compute the total abundance, while taking into account the variability in coverage probabilities, the posterior distributions of D 's may be treated as a function of environmental and geographic variables. Assuming the random placement of transect lines, density of the species may be modeled for the survey region, providing the total abundance. Such an approach, although not in the Bayesian framework, has been used for abundance estimation of dolphins in the Mediterranean (Cañadas and Hammond, 2006, 2008).

We have shown that the Bayesian approach is a powerful tool for line-transect analysis, especially when information on

ESW and density is available from other studies and new data are limited. Even with the additional sampling uncertainty to the analysis, the precisions of parameter estimates are comparable to or better than the equivalent analysis using Distance. The Bayesian approach is not a replacement for the Distance approach, however. Because analysis can be done relatively easily using computer packages, Distance and Bayesian approaches can be used as complementary analytical tools in a line-transect analysis. We think that the inference on parameters should be based on as much information as possible. The Bayesian and Distance approaches should be considered as two possible methods to make inference about the density and abundance of a population.

Acknowledgements

We thank Eric Archer, Megan Ferguson, and George Watters at the SWFSC for their constructive inputs. Two anonymous reviewers provided valuable inputs which improved the quality of the manuscript.

Appendix A. Derivation of the joint posterior on N_a and λ for unbinned data

The likelihood function for the perpendicular distance is:

$$f(y|\lambda) = 2\sqrt{\frac{\lambda}{2\pi}} \exp\left(-\frac{\lambda y^2}{2}\right) \quad (A1)$$

where $y \geq 0$. We assume that knowledge of λ can be represented by a gamma distribution with hyperparameters α and β , where $\lambda > 0$, $\alpha > 0$, and $\beta > 0$. The prior distribution on λ is thus

$$p(\lambda|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\lambda\beta) \quad (A2)$$

Using these models, detection probability (P_a) can be written as:

$$P_a = \frac{\int_0^W g(y|\lambda) dy}{W} = \frac{1}{W} \sqrt{\frac{2\pi}{\lambda}} \int_0^W \sqrt{\frac{\lambda}{2\pi}} \exp\left(-\frac{\lambda y^2}{2}\right) dy \quad (A3)$$

Using the properties of the normal distribution (A3) can be expressed with the standard normal cumulative distribution function ($\Phi(x)$):

$$P_a = \frac{1}{W} \sqrt{\frac{2\pi}{\lambda}} \left[\Phi\left(W\sqrt{\lambda}\right) - 0.5 \right] \quad (A4)$$

Given the total number of available objects in the sampled area (N_a), and the detection probability (P_a), the observed number of objects (n) follows binomial distribution:

$$p(n|N_a, P_a) = \binom{N_a}{n} P_a^n (1 - P_a)^{N_a-n} \quad (A5)$$

Substituting the binomial observation likelihood function, the half-normal detection likelihood function, and the prior distributions into the joint posterior distribution and using the conjugate property between normal and gamma distributions, the joint posterior of N_a and λ conditional on the observed data is:

$$p(N_a, \lambda | y, n) \propto \begin{cases} p(N_a)p(\lambda)p(n|N_a, \lambda)p(y|\lambda) \\ p(N_a)\lambda^\alpha \exp(-\lambda\beta) \frac{\Gamma(N_a+1)}{\Gamma(N_a-n+1)} P_a^n (1-P_a)^{N_a-n} \prod_{i=1}^n \left(\left(\frac{\lambda}{2\pi} \right)^{1/2} \exp\left(-\frac{\lambda y_i^2}{2}\right) \right) \\ p(N_a) \frac{\Gamma(N_a+1)}{\Gamma(N_a-n+1)} P_a^n (1-P_a)^{N_a-n} \lambda^\alpha \exp(-\lambda\beta) \left(\frac{\lambda}{2\pi} \right)^{n/2} \prod_{i=1}^n \left(\exp\left(-\frac{\lambda y_i^2}{2}\right) \right) \\ p(N_a) \frac{\Gamma(N_a+1)}{\Gamma(N_a-n+1)} P_a^n (1-P_a)^{N_a-n} \lambda^{\alpha+(n/2)} \left(\frac{1}{2\pi} \right)^{n/2} \exp(-\lambda\beta) \exp\left(-\frac{\lambda}{2} \sum_{i=1}^n y_i^2\right) \\ p(N_a) \frac{\Gamma(N_a+1)}{\Gamma(N_a-n+1)} P_a^n (1-P_a)^{N_a-n} \lambda^{\alpha+(n/2)} \exp\left(-\lambda \left(\beta + \frac{1}{2} \sum_{i=1}^n y_i^2 \right)\right) \\ p(N_a)\lambda^{\alpha_2-1} \exp(-\lambda\beta_2) \frac{\Gamma(N_a+1)}{\Gamma(N_a-n+1)} P_a^n (1-P_a)^{N_a-n} \end{cases} \quad (\text{A6})$$

where $\alpha_2 = \alpha + n/2$ and $\beta_2 = \beta + \sum_{i=1}^n y_i^2/2$, $i = 1, \dots, n$.

Appendix B. Prior distribution on effective strip width

For the prior distribution on λ , which is the precision of the half-normal detection function, we express the parameter in terms of the effective strip width (ESW). Following standard line-transect notation, ESW is denoted μ and defined as:

$$\mu = \frac{1}{f(0)} \quad (\text{A7})$$

where $f(0) = (2/\pi\sigma^2)^{1/2}$ for the half-normal detection function and σ^2 is the variance. Here, we redefine $f(0)$ in terms of ESW (μ). Solving for $1/\sigma^2$ yields

$$\frac{1}{\sigma^2} = \frac{\pi}{2\mu^2} = \lambda \quad (\text{A8})$$

where $\mu > 0$ and $\sigma^2 > 0$. Consequently, there is a one-to-one relationship between μ and σ^2 of the detection function.

To preserve the conjugate property between the gamma and half-normal distributions, the prior distribution on the ESW (μ) is transformed into a gamma prior distribution for λ . The assumption that $\lambda \sim \text{Gamma}(\alpha, \beta)$ results in $\sigma^2 \sim \text{inv-Gamma}(\alpha, \beta)$, and because $\sigma^2 = 2\mu^2/\pi$, $\mu^2 \sim \text{inv-Gamma}(\alpha, \beta_1)$, where $\beta_1 = (\pi/2)\beta$. Finally, we approximate the distribution of μ by another gamma distribution, i.e., $\mu \sim \text{Gamma}(\alpha_\mu, \beta_\mu)$. Although mathematically μ does not follow a gamma distribution when $\mu^2 \sim \text{inv-Gamma}(\alpha, \beta)$, the gamma distribution is convenient and reasonable for practical purposes (Appendix C).

During an analysis, the transformation process is reversed. From the information on μ , we compute the distribution of λ , which is approximated by a gamma distribution. We start with the prior information on the mean and variance of ESW ($\hat{\mu}$ and v_μ , respectively). We find appropriate gamma parameters $\alpha_\mu = \hat{\mu}^2/v_\mu$ and $\beta_\mu = \hat{\mu}/v_\mu$. We then generate random variables (μ) from this gamma distribution. Another gamma distribution then is fitted to the inverse of the squares ($1/\mu^2$), using the method of moments. Let the parameters of this fitted gamma distribution be α_0 and β_0 :

$$\frac{1}{\mu^2} = \delta \sim \text{Gamma}(\alpha_0, \beta_0). \quad (\text{A9})$$

Equivalently,

$$f(\delta|\alpha_0, \beta_0) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \delta^{\alpha_0-1} \exp(-\beta_0\delta),$$

where $\delta > 0$, $\alpha_0 > 0$, and $\beta_0 > 0$. Rewriting δ in terms of λ ,

$$\delta = \frac{2\lambda}{\pi},$$

because $\lambda = \pi/(2\mu^2)$ (A8). The distribution of λ can be found via transformation. Let $\lambda = \pi\delta/2$,

$$\begin{aligned} g(\lambda|\alpha_0, \beta_0) &= \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \left(\frac{2\lambda}{\pi} \right)^{\alpha_0-1} \exp\left(-\frac{2\beta_0\lambda}{\pi}\right) \left| \frac{d}{d\lambda} \frac{2\lambda}{\pi} \right| \\ &= \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \left(\frac{2}{\pi} \right)^{\alpha_0-1} \left(\frac{2}{\pi} \right) \lambda^{\alpha_0-1} \exp\left(-\frac{2\beta_0\lambda}{\pi}\right) \\ &= \frac{(2\beta_0/\pi)^{\alpha_0}}{\Gamma(\alpha_0)} \lambda^{\alpha_0-1} \exp\left(-\frac{2\beta_0\lambda}{\pi}\right), \end{aligned}$$

where $\lambda > 0$, $\alpha_0 > 0$, and $2\beta_0/\pi > 0$. This shows that $g(\lambda|\alpha_0, \beta_0)$ is another gamma distribution with parameters α_0 and $2\beta_0/\pi$.

Appendix C. Derivation of the exact distribution of μ , when $1/\mu^2 \sim \text{Gamma}(\alpha, \beta)$

Let $X = \mu^2$ and $Y = g(x) = \sqrt{x}$.

$$\begin{aligned} f(x) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{x} \right)^{\alpha-1} \exp\left(-\frac{\beta}{x}\right) \\ g^{-1}(y) &= y^2 \end{aligned}$$

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| = \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{y^2} \right)^{\alpha-1} \exp\left(-\frac{\beta}{y^2}\right) 2y \\ &= \frac{2\beta^\alpha}{\Gamma(\alpha)} y^{3-2\alpha} \exp\left(-\frac{\beta}{y^2}\right) \end{aligned} \quad (\text{A10})$$

where $y \geq 0$, $\alpha > 0$, $\beta > 0$. Although this probability density function is not a gamma distribution, we approximate this distribution with a gamma distribution. We think the difference between the two distributions is inconsequential for analysis (Fig. A1).

Appendix D. WinBugs model specification to conduct the proposed Bayesian analysis with unbinned data and the wooden stake dataset. Only one set of initial values is provided. Data are not truncated.

```

model
{
  for (i in 1:n){
    y[i] ~ dnorm(0, tau)
    n ~ dbin(p, N)
  }
  p <- (1/W) * sqrt(2*PI/tau) * (phi(W*sqrt(tau)) - 0.5)
  tau ~ dgamma(0.001,0.001)
  N ~ dunif(69, 4000)
  D <- N/(2*L*W)
  PI <- 3.14159
}

```

Initial values:
list(tau = 5, N = 200)

WOODEN STAKE DATA: note density is in unit of per hectare
list(n=68, L = 0.1000, W = 31.31, y = c(2.02, 0.45, 10.40, 3.61,
0.92, 1.00, 3.40, 2.90, 8.16, 6.47, 5.66, 2.95, 3.96, 0.09,
11.82, 14.23, 2.44, 1.61, 31.31, 6.50, 8.27, 4.85, 1.47, 18.60,
0.41, 0.40, 0.20, 11.59, 3.17, 7.10, 10.71, 3.86, 6.05, 6.42,
3.79, 15.24, 3.47, 3.05, 7.93, 18.15, 10.05, 4.41, 1.27, 13.72,
6.25, 3.59, 9.04, 7.68, 4.89, 9.10, 3.25, 8.49, 6.08, 0.40,
9.33, 0.53, 1.23, 1.67, 4.53, 3.12, 3.05, 6.60, 4.40, 4.97,
3.17, 7.67, 18.16, 4.08))

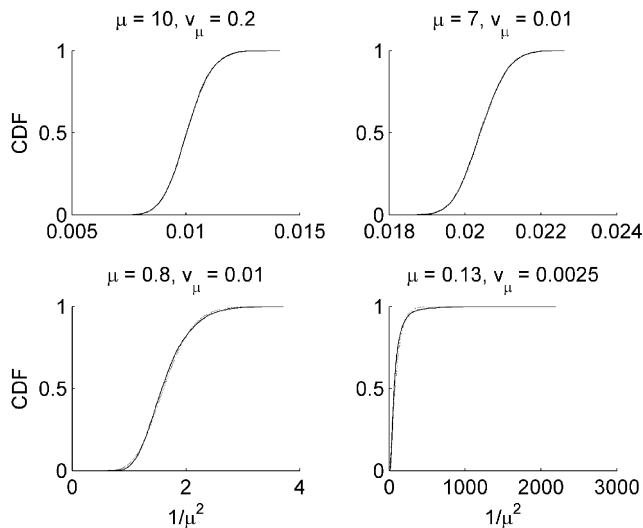


Fig. A1. Comparison of cumulative distributions (CDFs) of $1/\mu^2$ between the empirical (solid) and fitted gamma distributions (dotted). An empirical distribution was created by taking the squared inverse of gamma distributed random numbers with a mean ($\hat{\mu}$) and variance (v_{μ}). Another gamma distribution was fitted to the squared inverses using the method of moments (dotted line). In the analysis, gamma distribution (dotted) is used to approximate the true (solid) distribution.

References

- Berg, B.A., 2004. Markov Chain Monte Carlo simulations and Their Statistical Analysis. World Scientific, Singapore.
- Borchers, D.L., Burnham, K.P., 2004. General formulation for distance sampling. In: Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L.D., Borchers, L., Thomas, L. (Eds.), Advanced Distance Sampling. Oxford University Press, New York, NY, pp. 6–30.
- Buckland, S.T., 1985. Perpendicular distance models for line transect sampling. Biometrics 41, 177–195.
- Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., Thomas, L., 2001. Introduction to Distance Sampling. Oxford University Press, New York, NY.
- Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., Thomas, L., 2004. Advanced Distance Sampling. Oxford University Press, New York, NY.
- Burnham, K.P., Anderson, D.R., Laake, J.L., 1980. Estimation of density from line transect sampling of biological populations. Wildlife Monograph 72, supplement to Journal of Wildlife Management 44.
- Carlin, B.P., Clark, J.S., Gelfand, A.E., 2006. Elements of hierarchical Bayesian inference. In: Clark, J.S., Gelfand, A.E. (Eds.), Hierarchical Modeling for the Environmental Sciences. Oxford University Press, New York, NY, pp. 3–24.
- Casella, G., Berger, R.L., 2002. Statistical Inference, second edition. Duxbury, Pacific Grove, CA.
- Cañadas, A., Hammond, P.S., 2006. Model-based abundance estimates for bottlenose dolphins off southern Spain: implications for conservation and management. Journal of Cetacean Research 8, 13–27.
- Cañadas, A., Hammond, P.S., 2008. Abundance and habitat preferences of the short-beaked common dolphin *Delphinus delphis* in the southwestern Mediterranean: implications for conservation. Endangered Species Research 4, 309–331.
- Clark, J.S., 2007. Models for Ecological Data. Princeton University Press, Princeton, NJ.
- Clark, J.S., Bjørnstad, O.N., 2004. Population time series: process variability, observation errors, missing values, and hidden states. Ecology 85, 3140–3150.
- Clark, J.S., Ferraz, G., Oguge, N., Hays, H., DiCostanzo, J., 2005. Hierarchical Bayes for structured, variable populations: from recapture data to life-history prediction. Ecology 86, 2232–2244.
- Dennis, B., 1996. Discussion: should ecologists become Bayesians? Ecological Applications 6, 1095–1103.
- Dennis, B., 2004. Statistics and the scientific method in ecology. In: Taper, M.L., Lele, S.R. (Eds.), The Nature of Scientific Evidence. The University of Chicago Press, Chicago, IL, pp. 327–360.
- Ellison, A.M., 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. Ecological Applications 6, 1036–1046.
- Gelman, A., 2008a. Objections to Bayesian statistics. Bayesian Analysis 3, 445–450.
- Gelman, A., 2008b. Rejoinder. Bayesian Analysis 3, 467–478.
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis, second edition. Chapman and Hall/CRC, Boca Raton, FL.
- Gerrodette, T., Forcada, J., 2005. Non-recovery of two spotted and spinner dolphin populations in the eastern tropical Pacific Ocean. Marine Ecology Progress Series 291, 1–21.
- Gilbert, D.W., Anderson, D.R., Ringelman, J.K., Szymczak, M.R., 1996. Response of nesting ducks to habitat and management on the Monte Vista National Wildlife Refuge, Colorado. Wildlife Monographs 131.
- Goodman, D., 2004. Taking the prior seriously: Bayesian analysis without subjective probability. In: Taper, M.L., Lele, S.R. (Eds.), The Nature of Scientific Evidence. The University of Chicago Press, Chicago, IL, pp. 379–400.
- Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.
- Hogg, R.V., Craig, A.T., 1995. Introduction to Mathematical Statistics, fifth edition. Prentice-Hall, Inc., Upper Saddle River, NJ.
- Karunamuni, R.J., Quinn II, T.J., 1995. Bayesian estimation of animal abundance for line transect sampling. Biometrics 51, 1325–1337.
- Laake, J.L., 1978. Line transect estimators robust to animal movement. MS Thesis. Utah State University, Logan, UT, USA. 55 pp.

- Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS—a Bayesian modelling framework: concepts structure, and extensibility. *Statistics and Computing* 10, 325–337.
- Marques, F.F.C., Buckland, S.T., 2004. Covariate models for the detection function. In: Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L.D., Borchers, L., Thomas, L. (Eds.), *Advanced Distance Sampling*. Oxford University Press, New York, NY, pp. 31–47.
- Martinez, W.L., Martinez, A.R., 2002. *Computational Statistics Handbook with Matlab*. Chapman & Hall/CRC, Boca Raton, FL.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* 21, 1087–1092.
- Quang, P.X., 1990. Confidence intervals for densities in line transect sampling. *Biometrics* 46, 459–472.
- Quinn, T.J., Gallucci, V.F., 1980. Parametric models for line-transect estimators of abundance. *Ecology* 61, 293–302.
- Raftery, A.E., Schweder, T., 1993. Inference about the ratio of two parameters, with application to whale censusing. *The American Statistician* 47, 259–264.
- Seber, G.A.F., 1973. *The Estimation of Animal Abundance*. Macmillan Publishing Co, Inc., New York, NY.
- Sen, A.R., Torgny, J., Smith, G.E.J., 1974. On the line transect sampling method. *Biometrics* 30, 329–340.
- Strindberg, S., Buckland, S.T., Thomas, L., 2004. Design of distance sampling surveys and geographic information systems. In: Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L.D., Borchers, L., Thomas, L. (Eds.), *Advanced Distance Sampling*. Oxford University Press, New York, NY, pp. 190–228.
- Thomas, L., Laake, J.L., Strindberg, S., Marques, F.F.C., Buckland, S.T., Borchers, D.L., Anderson, D.R., Burnham, K.P., Hedley, S.L., Pollard, J.H., Bishop, J.R.B., 2003. Distance 4.1. Release 2. Research Unit for Wildlife Population Assessment, University of St. Andrews, JK. <http://www.ruwpa.st-and.ac.uk/distance/>.
- Thomas, L., Burnham, K.P., Buckland, S.T., 2004. Temporal inferences from distance sampling surveys. In: Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L.D., Borchers, L., Thomas, L. (Eds.), *Advanced Distance Sampling*. Oxford University Press, New York, NY, pp. 71–107.
- Wade, P.R., 2000. Bayesian methods in conservation biology. *Conservation Biology* 14, 1308–1316.